Letter to the Editor,

Somatic mutations of cell-free circulating DNA detected by Next Generation Sequencing reflect the genetic changes in both Germinal Center B-Cell like and Activated B-Cell like Diffuse Large B-Cell Lymphoma tumors at the time of diagnosis.

Elodie Bohers (1), Pierre Julien Viailly (1), Sydney Dubois (1) , Philippe Bertrand (1), Catherine Maingonnat (1), Sylvain Mareschal (1), Philippe Ruminy (1), Jean-Michel Picquenot (2) Christian Bastard (1), Fabienne Desmots (3), Thierry Fest (3), Karen Leroy (4), Hervé Tilly (1,5) and Fabrice Jardin (1,5).

(1) INSERM U918, Centre Henri Becquerel, Université de Rouen, IRIB, Rouen, France;

(2) Department of Pathology, Centre Henri Becquerel, Rouen, France

(3) UMR INSERM U917, CHU Pontchaillou, Rennes, France

(4) INSERM U955, Henri Mondor Hospital, Creteil, France

(5) Department of Clinical Hematology, Centre Henri Becquerel, Rouen, France

Word count: 1880

Key words: diffuse large B-cell lymphoma, NGS, liquid biopsy, mutations

Number of figure: 1

Number of Table: 1

Number of supplementary files: 1

Reference Number: 15

Running title: Somatic mutations of cell-free circulating DNA in DLBCL

Corresponding Author: Fabrice Jardin, MD-PhD; email: fabrice.jardin@chb.unicancer.fr; tel +33 (0)232082909

Diffuse Large B-cell Lymphoma (DLBCL) is the most common form of lymphoma, accounting for 30-40% of newly diagnosed Non-Hodgkin Lymphomas (NHL). The molecular heterogeneity of DLBCLs has been deciphered by gene expression profiling, and DLBCLs have been divided into three main molecular subtypes: the Germinal Center B-cell like (GCB) subtype, the Activated B-Cell like (ABC) subtype, and the Primary Mediastinal B-cell Lymphoma (PMBL) subtype with distinct clinical outcomes and responses to immunochemotherapy. Next Generation Sequencing (NGS) technologies, which allow for massively parallel, high-throughput DNA sequencing, have emerged over the past decade and have provided new insights into the genomic characterization of DLBCL. Recurrent Single Nucleotide Variants (SNVs) are now well defined and provide new therapeutic opportunities for the three molecular subtypes. These SNVs target genes that play a crucial role in several pathways including BCR signaling (*CD79A/CD79B*), NFκ-B (*CARD11*), Toll-like receptor signaling (*MYD88*), immunity (*CD58, TNFSRF14, B2M*), cell cycle/apoptosis (*TP53, BCL2*) or epigenetic regulation (*EZH2,CREBBP, MLL2*)(1, 2) .

Recently, whole exome sequencing in breast cancer has shown that mutations observed in the tumor could also be detected in circulating cell-free DNA (cfDNA) and could be used to detect genetic changes during treatments and relapse, defining the concept of "liquid biopsy"(3). In DLBCL, whereas tumoral circulating cells or leukemic phase are not usually detectable, clonotypic sequences have been constantly detected in cfDNA extracted from serum/plasma or PBMC(4-8).

In this study we sought to determine by routinely applicable NGS technology whether the pattern of acquired SNVs observed in tumor DNA could also be detected in cfDNA in DLBCL patients at the time of diagnosis. For this purpose, we analyzed twelve DLBCL cases with available matched tumor DNA and plasma collected at the time of diagnosis. Patients harboring typical GCB/ABC-related mutations targeting CD79A*/B, EZH2, CARD11* or *MYD88* genes previously identified by Sanger Method were selected(9). This study was approved by the regional ethical committee (numbered as CPP N°01/006/2014).

The main clinical features of the patients are summarized in **Table 1**. None of the selected cases harbored detectable circulating lymphoma cells by routine blood smear examination. Of note, no peripheral blood cytometry was performed, in accordance with our center's initial staging procedures for DLBCL patients. Tumor DNA was extracted from frozen lymph node samples by standard methods. cfDNA was extracted from archived EDTA-anticoagulated plasma aliquots (1mL) stored at -80°C using the QIAamp® Circulating Nucleic Acid Kit (Qiagen) (with the QIAvac 24 Plus vacuum manifold, following the manufacturer's instructions), and concentrations were measured using Fluorometric assay (Qubit® dsDNA HS Assay Kit, Life Technologies). The mean cfDNA concentration in plasma was 1.65 ng/µl (range 0.46-11.2) (**Table 1**). The Cell of origin (COO) signature was determined by cDNA-mediated annealing, Selection, extension, and Ligation (DASL) technology based on the expression of 19 genes as previously reported(10). Among the 12 cases analyzed, 5 belonged to the ABC subgroup, 6 to the GCB subgroup and one case was unclassified (Table 1).

Sequencing of tumor DNA was performed using Ion Torrent Personal Genome Machine (PGM, Life technologies). Ten ng of genomic DNA were submitted to NGS using a laboratory-developed Lymphopanel set, designed to identify mutations in 34 genes relevant to lymphomagenesis (**Supplementary file 1A**). This design covers 87 703 bases and generates 872 amplicons. Amplified

libraries (Ion AmpliSeq™ Library Kit 2.0) were submitted to emulsion PCR with the Ion OneTouch™ 200 Template Kit (Life Technologies) using the Ion OneTouch™ System (Life Technologies) according to the manufacturer's instructions. The templated Ion Sphere™ Particles (ISPs) were enriched with the Ion OneTouch™ Enrichment System and loaded and sequenced on an Ion 316™ v2 Chip (Life Technologies).

After alignment to a reference genome sequence (hg19) and variant calling procedure, variants were filtered through a dedicated bioinformatic pipeline that eliminated synonymous variants and variants with a Variant Allele Frequency greater than 1% in the 1000 genome database (considered as polymorphisms). Only non-synonymous SNV/In/Del with a quality score > 22 and/or confirmed by Sanger experiment were retained as acquired somatic mutations in lymphoma cases and were used for the subsequent sequencing of the matched cfDNA (for the detailed pipeline see **supplementary file 1B**). Somatic mutations identified in tumor DNA were used to build a hot-spot file enabling higher sensibility tracking of somatic mutations in circulating DNA.

In order to increase variant detection sensitivity and specificity in cfDNA, we exclusively amplified amplicons targeting mutations detected in the corresponding tumor DNA by the complete Lymphopanel by performing a dedicated sequencing procedure with a pool of oligonucleotide primers selected among the 872 pairs provided by Life Technologies. The procedure to make librairies and sequence amplicons was the same as for tumor DNA but used a 314™ v2 Chip. When possible, circulating DNA was extracted and sequenced from 2 different aliquots of plasma.

The SNV/In/Del detected in both tumor and circulating DNA are indicated in **Table 1** with their corresponding variant allele frequency (VAF). As expected, we identified a typical SNV pattern in the five ABC DLBCL cases including mutations targeting *MYD88, CD79A/B, PIM1, PRDM1, CARD11* or *IRF4*, whereas *EZH2, BCL2, GNA13,* or *TNFSRF14* were mutated in the unclassified and GCB DLBCL cases. *MLL2 (KMT2D), CREBBP* or *ITPKB* were targeted by somatic mutations shared in the two COO subtypes. The sequencing depths obtained for each sample and targets are indicated in Table 1.

The mean number of reads targeting the mutated regions for tumor DNA was 241 (range 23-741) as compared to 5,987 (range 6-22,541) for plasma DNA, indicating a 24-fold mean depth sequencing increase. The mean VAF in the tumor DNA was 35% (range 17-64%) as compared to a mean of 11% for plasma DNA (range 2-89%) (**Table 1**).

In 11/12 DLBCL cases, we observed somatic mutations in cfDNA, similar or partially similar to those observed in the tumor (**Figure 1, supplementary figure 1 and table 1**). We defined the concordance rate as the ratio of the number of mutated cfDNA genes and the number of mutated tumor DNA genes. This rate ranges from 33% to 100% (5 cases) in the 11/12 cases with detected mutated cfDNA. Overall, the median concordance rate between tumor DNA and cfDNA is 85%.

In one case (#1251), SNVs observed in the tumor DNA were not detected in plasma DNA. In an additional case (#1559), SNVs were barely detectable (VAF of 0.5% for two variants). Of note, both cases displayed a limited disease (stage I or II) and normal LDH level, indicating that tumor specific circulating cfDNA amount is at least partially related to tumor burden. Despite a low amount of circulating DNA extracted from plasma for cases #1251 and #1559, we obtained adequate sequencing quality and depth (overall number of reads sequenced with mutated target = 4,685 and

51,195 respectively, Table 1), indicating that in some rare cases, the tumor specific cfDNA is absent or beneath the level of sensitivity of the NGS sequencing method used. Of note, in case #1631 characterized by a limited stage I disease, SNVs were detected with a mean VAF of 5.2% in plasma DNA, as compared to a mean VAF of 34.6% in the tumor DNA (**Table 1**). In contrast, cases #1639 and #1768 (both with stage IV disease and elevated LDH level) displayed a high proportion of tumor specific circulating DNA as indicated by the high VAF observed (**Table 1**). Interestingly, in these two cases, the sub-clonal distribution of certain mutations, as indicated by the VAF distribution of each individual variant, was also observed in cfDNA, suggesting that sequencing cfDNA can reflect the SNV pattern observed in tumor cells in some instances (**Figure 1** and **supplementary figure 1**). In case # 1524, despite a sufficient number of relevant reads (> 4,000) we failed to detect the B2M SNV present in the lymph node biopsy. This result was confirmed by manual Integrative Genomics Viewer (IGV) checking, suggesting that the B2M SNV is present only in a subclone caught in the biopsy sample but not highlighted by the cfDNA that reflects the entire tumor burden.

By contrast, in some cases, the number of target reads is clearly insufficient for adequate SNV detection (# case 964, CARD11; case 1003, MYD88 see table 1), most likely reflecting the low amount of cfDNA available rather than a true clonal divergence between tumor DNA and cfDNA. This was not observed in cases with higher cfDNA amount.

Failure to detect SNVs in cfDNA appears related more to the proportion of tumor specific DNA quantity than to the total amount or quality of total cfDNA. Of note, we failed to find any tumor-related SNV in DNA extracted from 11/12 PBMC using this approach (data not shown), indicating that serum or plasma is preferable to detect mutated circulating DNA. In a large cohort of Hodgkin lymphoma, mantle cell lymphoma and DLBCL, increased levels of plasma DNA (determined using quantitative PCR for the β-globin gene) were associated with advanced stage disease, presence of B-symptoms, elevated lactate dehydrogenase levels, and age > 60 years also indicating that the amount of circulating DNA is partially related to tumor burden(8). Furthermore, it has been shown in a cohort of EBV-positive lymphoma that serum and plasma were equivalent to detect Lymphoma-specific DNA but that only the lymphoma specific DNA could be used to monitor disease response in lymphoma(7).

To our knowledge this is the first report of the detection of non-immunoglobulin somatic mutations in DLBCL from circulating DNA by routine NGS sequencing, enabling the identification of lymphoma-specific cfDNA. Other quantitative approaches, including digital PCR, are also suitable and could be used in this setting for detecting recurrent translocations or mutations(11). More recently the LymphoSIGHT®, a high-throughput DNA sequencing method, was developed to detect and quantify circulating tumor DNA as minimal residual disease (MRD) and was able to predict both early treatment failure and relapse in newly diagnosed DLBCL patients, chronic lymphocytic leukemia or acute lymphoblastic leukemia(12-15). This approach is based on tumor DNA amplification using locus-specific primer sets for the immunoglobulin heavy/light-chain which failed in a substantial number of DLBCL cases. Importantly the Lymphopanel used in this study is able to detect at least one acquired SNV in 95% of DLBCL cases at initial diagnosis (manuscript in preparation) and may therefore constitute a simple routinely applicable test to provide the COO subtype or to detect targetable mutations at the time of diagnosis or relapse. However, its capacity to detect MRD with a high level of sensitivity remains to be determined and we can hypothesize that at least 5 to 10% of

DLBCL cases will not display any SNVs detectable by our Lymphopanel. Furthermore, cfDNA sequencing was successfully performed using the entire Lymphopanel (including the 34 targeted genes) in one case (data not shown), indicating that this approach is feasible without the knowledge of the tumor variant calling. However this requires an increase in the sequencing depth capacity and entails a substantial cost increase.

To conclude, our results indicate that cfDNA can also be used in DLBCL to detect somatic variants, validating the concept of "liquid biopsy" in this type of tumor(3). These preliminary results have prompted us to start a prospective study in the aim of serial sequencing cfDNA during DLBCL treatment and follow-up (registered on clinicaltrials.gov as NCT02339805). If these preliminary results are confirmed by a prospective study, new strategies should be proposed for both diagnosis and treatment tailoring based on the simple detection and quantification of SNV in plasma.

Conflicts of interest: The authors have no conflicts of interest to declare

Contributions: EB and CM performed PGM and Sanger sequencing experiments; FJ and EB designed the study; FJ, HT and CB supervised the experiments and provided clinical data; PJV, SM and PB performed bioinformatics and designed pipeline analysis; JMP performed histopathological review; PR performed Sanger analysis; FJ, KL, TF, EB, PB and PR designed the lymphopanel. All authors were involved in the manuscript writing and approved the final manuscript.

References

1. Morin RD, Gascoyne RD. Newly identified mechanisms in B-cell non-Hodgkin lymphomas uncovered by next-generation sequencing. Seminars in hematology. 2013 Oct;50(4):303-313.
2. Roschewski M, Staudt LM, Wilson WH. Diffuse large B-cell lymphoma-treatment approaches in the molecular era. Nature reviews Clinical oncology. 2014 Jan;11(1):12-23.
3. Diaz LA, Jr., Bardelli A. Liquid biopsies: genotyping circulating tumor DNA. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2014 Feb 20;32(6):579-586.
4. He J, Wu J, Jiao Y, et al. IgH gene rearrangements as plasma biomarkers in Non- Hodgkin's lymphoma patients. Oncotarget. 2011 Mar;2(3):178-185.
5. Hosny G, Farahat N, Hainaut P. TP53 mutations in circulating free DNA from Egyptian patients with non-Hodgkin's lymphoma. Cancer letters. 2009 Mar 18;275(2):234-239.
6. Mussolin L, Burnelli R, Pillon M, et al. Plasma cell-free DNA in paediatric lymphomas. Journal of Cancer. 2013;4(4):323-329.
7. Jones K, Nourse JP, Keane C, et al. Tumor-specific but not nonspecific cell-free circulating DNA can be used to monitor disease response in lymphoma. American journal of hematology. 2012 Mar;87(3):258-265.
8. Hohaus S, Giachelia M, Massini G, et al. Cell-free circulating DNA in Hodgkin's and non-Hodgkin's lymphomas. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO. 2009 Aug;20(8):1408-1413.

9. Bohers E, Mareschal S, Bouzelfen A, et al. Targetable activating mutations are very frequent in GCB and ABC diffuse large B-cell lymphoma. Genes, chromosomes & cancer. 2014 Feb;53(2):144-153.
10. Lanic H, Mareschal S, Mechken F, et al. Interim positron emission tomography scan associated with international prognostic index and germinal center B cell-like signature as prognostic index in diffuse large B-cell lymphoma. Leukemia & lymphoma. 2012 Jan;53(1):34-42.
11. Shuga J, Zeng Y, Novak R, et al. Single molecule quantitation and sequencing of rare translocations using microfluidic nested digital PCR. Nucleic acids research. 2013 Sep;41(16):e159.
12. Armand P, Oki Y, Neuberg DS, et al. Detection of circulating tumour DNA in patients with aggressive B-cell non-Hodgkin lymphoma. British journal of haematology. 2013 Oct;163(1):123-126.
13. Logan AC, Vashi N, Faham M, et al. Immunoglobulin and T cell receptor gene high-throughput sequencing quantifies minimal residual disease in acute lymphoblastic leukemia and predicts post-transplantation relapse and survival. Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation. 2014 Sep;20(9):1307-1313.
14. Logan AC, Zhang B, Narasimhan B, et al. Minimal residual disease quantification using consensus primers and high-throughput IGH sequencing predicts post-transplant relapse in chronic lymphocytic leukemia. Leukemia. 2013 Aug;27(8):1659-1665.
15. Roschewski M, Dunleavy K, Pittaluga S, et al. Monitoring of Circulating Tumor DNA As Minimal Residual Disease in Diffuse Large B-Cell Lymphoma. Blood. 2014;124(21):139 Abstr.

									Tumor DNA			Circulating DNA				
UPN	Sex	Age	Stage	IPI	LDH (xUNL)	Bone marrow involvement	Phenotype	Gene	VAF (%)	read number	mean VAF (%)	VAF (%)	read number	mean VAF (%)	Allele Call	concentration (ng/µl)
								ITPKB (1)	40.9	79/193		2.5	158/6313		Heterozygous	
								ITPKB (2)	40.6	78/192		2.4	152/6260		Heterozygous	
964	F	17	IV	3	5	0	GCB	*CARD11*	17.1	13/76	40	0.0	0/54	2	Absent	2.02
								EZH2	26.3	76/289		2.8	126/4435		Heterozygous	
								B2M	75.6	102/135		4.0	130/4016		Heterozygous	
								ITPKB	35.3	36/102		1.4	22/1540		Heterozygous	
1003	F	75	III	3	2.5	0	ABC	*MYD88*	26.5	67/253	40	0.0	0/21	5	Absent	1.94
								CIITA	51.9	14/27		16.7	1/6		Heterozygous	
								CD79B	46.3	93/201		2.1	369/17260		Heterozygous	
								TNFRSF14	52.5	74/141		0.0	0/1113		Absent	
1251	M	#	IE	0	0.8	0	GCB	*EZH2*	9.8	38/389	31	0.0	0/2519	0	Absent	0.46
								SOCS1	30.4	7/23		0.0	0/1053		Absent	
1437	M	76	III	4	5.85	0	ABC	*CD79A (1)*	17.0	53/318	17	89.1	344/386	89	Heterozygous	11.2
								CD79A (2)	17.0	53/318		89.1	344/386		Heterozygous	
								TNFRSF14	45.3	77/170		17.6	539/3064		Heterozygous	
								EZH2	22.3	106/475		7.0	818/11751		Heterozygous	
								KMT2D (1)	24.1	64/266		12.1	1269/10509		Heterozygous	
								KMT2D (2)	30.6	33/108		14.9	1015/6818		Heterozygous	
1524	F	45	III	2	1.52	0	GCB	*B2M*	33.3	44/132	34	0.0	0/4555	14	Absent	1.48
								CREBBP	34.4	45/131		14.2	47/331		Heterozygous	
								GNA13 (1)	43.7	52/119		23.1	1909/8254		Heterozygous	
								GNA13 (2)	43.7	52/120		23.1	1920/8314		Heterozygous	
								GNA13 (3)	44.2	53/120		23.4	1946/8312		Heterozygous	
								BCL2	17.1	42/245		4.8	409/8558		Heterozygous	
								ITPKB (1)	9.3	14/151		0.0	0/1713		Absent	
								ITPKB (2)	15.2	19/125		0.0	0/3582		Absent	
								ITPKB (3)	15.2	19/125		0.0	0/3577		Absent	
								ITPKB (4)	15.5	19/123		0.0	0/3570		Absent	
								MYD88	37.8	74/196		8.5	161/1888		Heterozygous	
								TNFAIP3	45.7	32/70		5.6	177/3152		Heterozygous	
								EZH2	5.9	44/741		2.1	152/7401		Heterozygous	
								KMT2D	36.1	122/338		5.7	207/3620		Heterozygous	
1528	F	#	III	4	1.39	0	GCB	*B2M*	50.0	63/126	31	3.2	78/2465	3	Heterozygous	1.94
								CIITA (1)	39.0	30/77		0.0	0/2130		Absent	
								CIITA (2)	7.7	14/183		0.0	0/9150		Absent	
								CIITA (3)	7.6	14/184		0.0	0/9162		Absent	
								GNA13 (1)	71.0	105/148		0.0	0/2695		Absent	
								GNA13 (2)	71.0	105/148		3.1	82/2672		Heterozygous	
								BCL2(1)	34.1	47/138		4.3	214/4929		Heterozygous	
								BCL2 (2)	32.9	45/137		4.2	209/5013		Heterozygous	
								MEF2B	33.3	137/411		5.9	654/11153		Heterozygous	
								CD58	92.9	92/99		0.0	0/13231		Absent	
								MYD88	51.8	187/361		0.0	0/11427		Absent	
1559	F	#	IE	1	0.8	0	ABC	*CREBBP*	46.3	154/333	64	0.0	0/5663	<1	Absent	0.618
								TP53	89.0	105/118		0.0	0/13012		Absent	
								CD79A (1)	52.0	122/235		0.5	20/3930		Absent	
								CD79A (2)	52.0	122/235		0.5	20/3932		Absent	
								TNFRSF14	13.8	44/318		2.1	6/284		Heterozygous	
								EZH2	13.8	41/298		0.0	0/6575		Absent	
1586	F	#	III	2	0.85	0	NA	*STAT6*	18.9	83/439	20	0.0	2/5517	<1	Absent	1.31
								CREBBP	33.6	49/146		1.3	65/4894		Absent	
								GNA13	17.7	11/62		0.0	0/4391		Absent	
								TNFRSF14	34.8	146/419		19.3	694/3587		Heterozygous	
								EZH2	9.9	57/575		1.8	62/3480		Heterozygous	
								STAT6	15.4	79/514		5.1	200/3889		Heterozygous	
1623	M	53	IV	2	1.03	0	GCB	*CREBBP (1)*	23.2	166/717	18	11.6	489/4222	11	Heterozygous	0.95
								CREBBP (2)	14.0	54/387		10.1	328/3239		Heterozygous	
								CD79B	17.1	83/486		18.0	839/4651		Heterozygous	
								MEF2B	11.8	74/626		9.9	1283/12969		Heterozygous	

Table 1 (to be continued)

									Tumor DNA			Circulating DNA				
UPN	Sex	Age	Stage	IPI	LDH (xULN)	Bone marrow involvement	Phenotype	Gene	VAF (%)	read number	mean VAF (%)	VAF (%)	read number	mean VAF (%)	Allele Call	concentration (ng/µl)
								MYD88	49.8	220/442		11.7	655/5586		Heterozygous	
								PIM1	20.6	14/68		2.7	130/4797		Heterozygous	
								CARD11 (1)	47.1	99/210		5.1	154/3028		Heterozygous	
1631	M	#	I	2	1.54	0	ABC	*CARD11 (2)*	47.9	68/142	35	5.5	344/6300	5	Heterozygous	1.1
								STAT6	18.1	96/532		4.3	570/13250		Heterozygous	
								TP53	32.4	72/222		3.9	522/13317		Heterozygous	
								CD79B	26.7	52/195		3.3	386/11853		Heterozygous	
								TNFRSF14	75.8	47/62		82.3	2400/2915		Heterozygous	
								MYD88	27.5	133/483		7.1	1616/22541		Heterozygous	
								EZH2	24.0	126/526		32.8	3259/9939		Heterozygous	
								KMT2D (1)	33.8	187/554		35.0	5878/16784		Heterozygous	
1639	M	#	IV	2	2.4	0	GCB	*KMT2D (2)*	28.3	13/46	39	22.9	671/2929	34	Heterozygous	10.2
								CREBBP	37.0	17/46		30.6	2918/9530		Heterozygous	
								BCL2 (1)	45.8	11/24		31.5	223/709		Heterozygous	
								BCL2 (2)	53.0	98/185		28.2	2267/8032		Heterozygous	
								EP300	29.6	34/115		38.3	4597/11990		Heterozygous	
								MYD88	62.1	216/348		45.0	5936/13194		Heterozygous	
								IRF4	28.5	75/263		21.2	929/4374		Heterozygous	
								PIM1 (1)	61.7	145/235		38.2	1541/4030		Heterozygous	
1768	M	#	IV	3	4.2	0	ABC	*PIM1 (2)*	29.8	34/114	52	25.9	987/3812	34	Heterozygous	1.82
								PRDM1(1)	81.9	149/182		34.2	1273/3724		Heterozygous	
								PRDM1(2)	80.7	221/274		43.4	5076/11689		Heterozygous	
								MYC	36.4	102/280		26.8	262/979		Heterozygous	
								CD79B	36.5	57/156		33.6	3041/9060		Heterozygous	

Table 1. Clinical characteristics and list of somatic variants (insertion/deletion/ single nucleotide variant) detected by sequencing in tumor DNA and cell-free plasma circulating DNA. Details of the locations of the mutations are indicated in the **Supplementary table 1**.

ABC: activated B-cell like; GCB: germinal center B-cell like; In/del: insertion/deletion; LDH: Lactate dehydrogenase; IPI: international prognosis index; SNV: single nucleotide variant; UPN: unique personal number; ULN: upper limit value; VAF: variant allele frequency

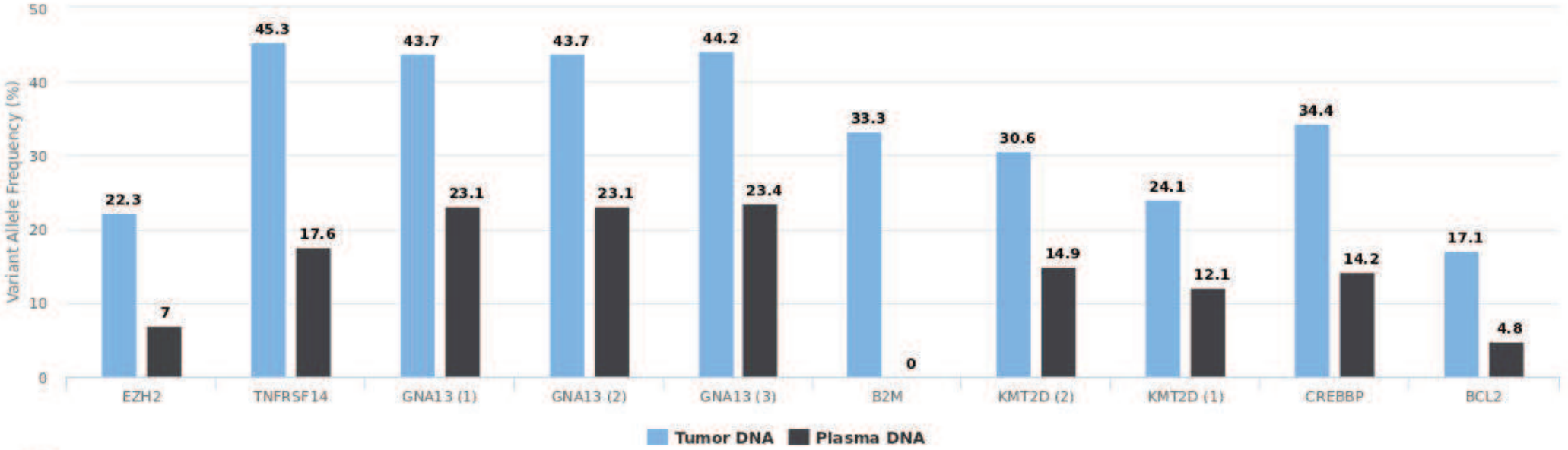
Figure legends

Figure 1. Representative examples of Variant allele frequency observed in tumor DNA and matched circulating cell-free DNA at the time of diagnosis.

A

#UPN1524

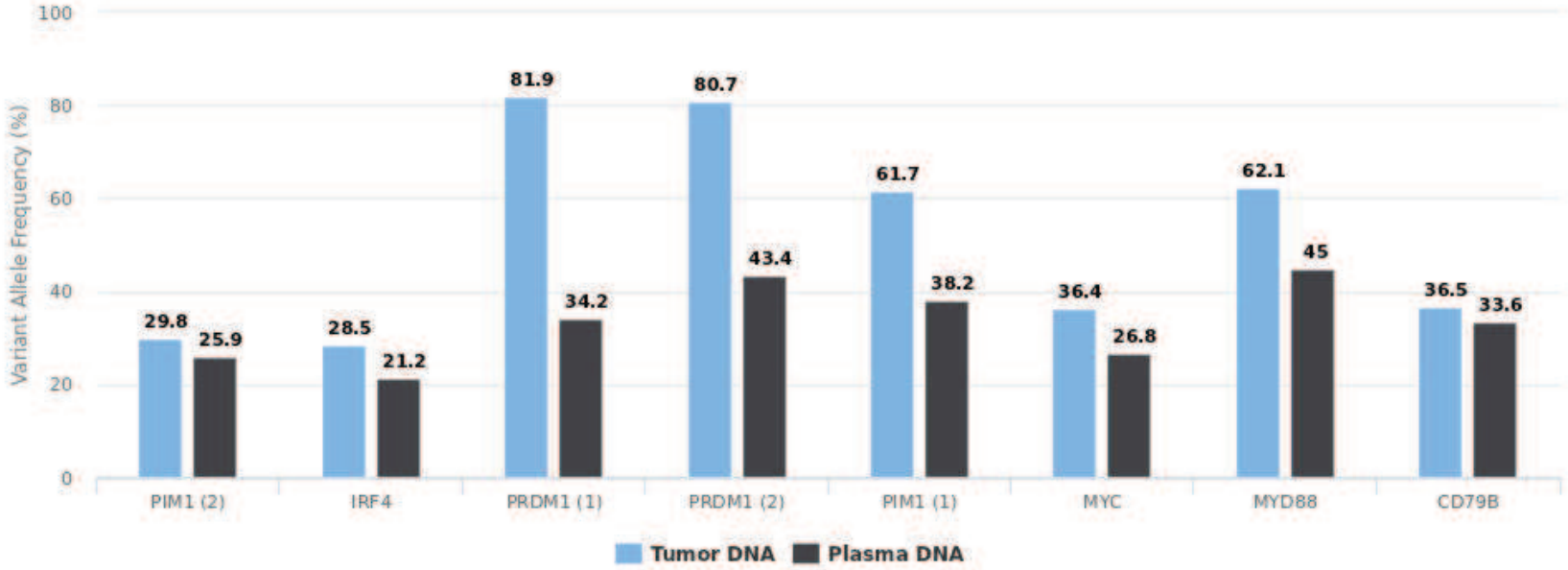
Relevant reads sequenced
Tumor: 1,886
Plasma: 70,466



B

#UPN1768

Relevant reads sequenced
Tumor: 1,852
Plasma: 50,862



Supplementary file 1. Lymphopanel and PGM data analysis

A. Lymphopanel set used for NGS experiments

Gene	Transcript Reference	Hotspot / Exons partially sequenced	Chromosomal location	Size sequenced (bp)
B2M	NM_004048	Hotspots exons 1 & 2, exon 3	15q21.1	360
BCL2	NM_000633	Hotspot exon 2	18q21.33	585
BRAF	NM_004333	Exon 15	7q34	119
CARD11	NM_032415	Coiled-coil domain exons 4-9	7p22.2	1121
CD58	NM_001779	Exons 1-6	1p13.1	753
CD79A	NM_001783	ITAM domain exons 4 & 5	19q13.2	183
CD79B	NM_000626	ITAM domain exons 5 & 6	17q23.3	141
CDKN2A	NM_000077+NM_058195+ NM_058197+NM_001195132	Exons 1, 2A, 2B, 3, 4 & 5	9p21.3	1,737
CDKN2B	NM_004936 + NM_0078487	Exons 1A, 1B & 2	9p21.3	1289
CIITA	NM_000246	Exons 1-19	16p13.13	3393
CREBBP	NM_004380	Exons 1-31	16p13.3	7323
EP300	NM_001429	Exons 1-31	22q13.2	7245
EZH2	NM_004456	SET domain, hotspots exon 16 & 18	7q36.1	177
FOXO1	NM_002015	Hotspots exon 1 & FH domain exon 2	13q14.11	780
GNA13	NM_006572	Exons 1-4	17q24.1	1134
ID3	NM_002167	Exons 1 & 2	1p36.12	360
IRF4/MUM1	NM_002460	Exons 2-9	6p25.3	1356
ITPKB	NM_002221	Exons 2-8	1q42.12	2830
KMT2D/MLL2	NM_003482	Exons 1-54	12q13.12	16614
MEF2B	NM_001145785	Exons 2-9	19p13.11	1107
MFHAS1	NM_004225	Exons 1-3	8p23.1	3159
MYC	NM_002467	Exons 1-3	8q24.21	1365
MYD88	NM_001172567	Exons 2-5	3p22.2	587
NOTCH1	NM_017617	PEST domain exon 34	9q34.3	1488
NOTCH2	NM_024408	Exons 26-28 & 34 (HD/PEST domains)	1p12-p11.2	2091
PIM1	NM_002648	Exons 1-6	6p21.2	942
PRDM1/BLIMP1	NM_001198	Exons 1-7	6q21	2478
SOCS1	NM_003745	Exon 2	16p13.13	636
STAT6	NM_001178078	Exons 9-14 (DNA binding domain hotspot)	12q13.3	795
TCF3	NM_001136139	B-HLH domain of E47 isoform exons 17 & 18	19p13.3	370
TNFAIP3	NM_006290	Exons 2-9	6q23.3	2373
TNFRSF14	NM_003820	Exons 1-8	1p36.32	852
TP53	NM_000546	Mutation hotspots exons 4-10	17p13.1	1004
XPO1	NM_003400	Exons 15-18	2p15	640

B. PGM Data analysis

Torrent Suite™ version 4.0 (Life Technologies) software was used to perform primary analysis, including signal processing, base calling, sequence alignment to the reference genome (hg19) and generation of Binary Alignment/Map (BAM) files. BAM files were used by Torrent Suite™'s Variant Caller to detect point mutations as well as short insertions and deletions using the "Somatic" and "Low Stringency" default parameters. VCF files generated by Variant Caller were annotated by ANNOVAR(1).

Data generated from tumor DNA samples were considered of sufficient quality when more than 90% of targeted bases were read at least 20 times with sequencing and mapping precisions of at least Q20. Only frameshift deletions and insertions, nonframeshift deletions and substitutions, splicing, nonsynonymous, stopgain or stoploss Single Nucleotide Variations (SNVs) were kept. Variants with a minimal Variant Allele Frequency (MAF) greater than 1% in the 1000 genome database were considered as polymorphisms and were discarded(2). A normal probability plot defined thresholds separating true positives [confirmed by Sanger sequencing, TVC (Torrent variant calling) quality score ≥ 22] from true negatives (discredited by Sanger sequencing, TVC score < 9.5) and highlighted a gray zone (9.5 ≤ TVC score < 22) in which variants must be confirmed by Sanger sequencing or pyrosequencing..

Further verification by Sanger sequencing was performed using a BigDye® Terminator v3.1 Cycle Sequencing Kit (Life Technologies) and an ABI PRISM 3130 analyzer (Life Technologies). Further verification by pyrosequencing was performed using the PyroMark PCR kit (Qiagen, France) with internal and sequencing primers designed using PyroMark software (Qiagen).

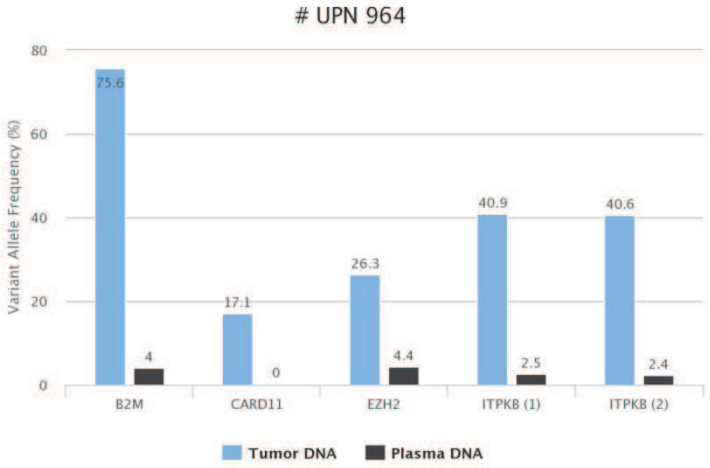
For circulating DNA variant calling, the parameter file was modified as follows: hotspot_min_variant_score: 3; hotspot_strand_bias: 1; downsample_to_coverage: 30 000. In addition, a dedicated hotspot VCF file generated from variants found in tumor DNA was used. This file instructs the Variant Caller to include these positions in its output files, including evidence for a variant and the filtering thresholds that disqualify a variant candidate.

References

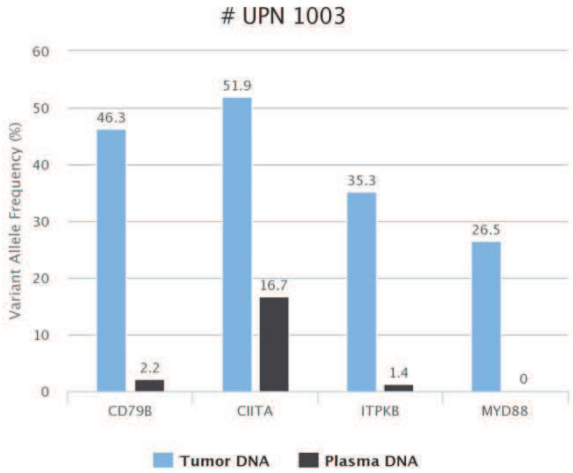
1. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic acids research. 2010 Sep;38(16):e164.
2. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012 Nov 1;491(7422):56-65.

Figure S1

Relevant reads sequenced
Tumor: 885
Plasma: 21,078



Relevant reads sequenced
Tumor: 582
Plasma: 18,827



Relevant reads sequenced
Tumor: 553
Plasma: 4,685

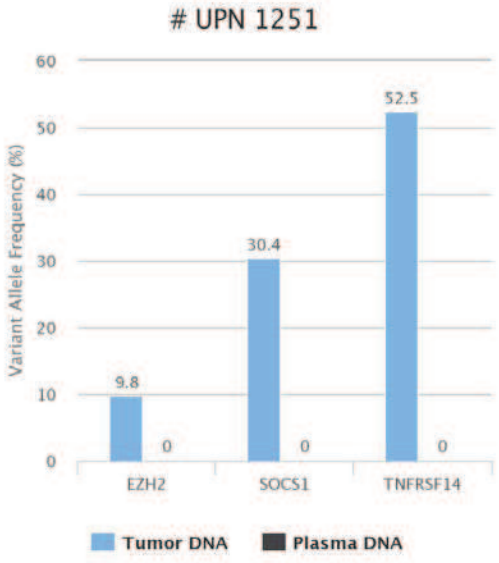
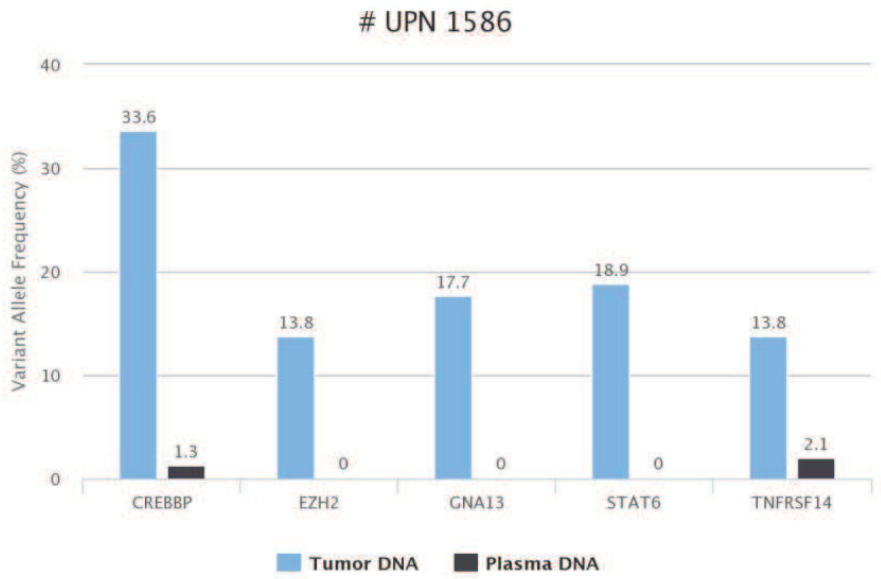


Figure S1

Relevant reads sequenced
Tumor: 1,263
Plasma: 21,661



Relevant reads sequenced
Tumor: 1,381
Plasma: 51,195

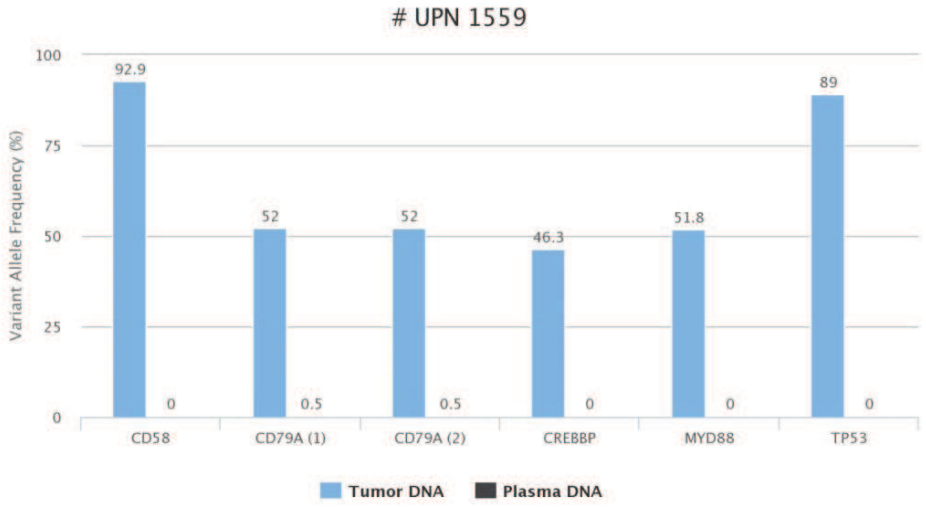


Figure S1

Relevant reads sequence
Tumor: 1,886
Plasma: 70,466

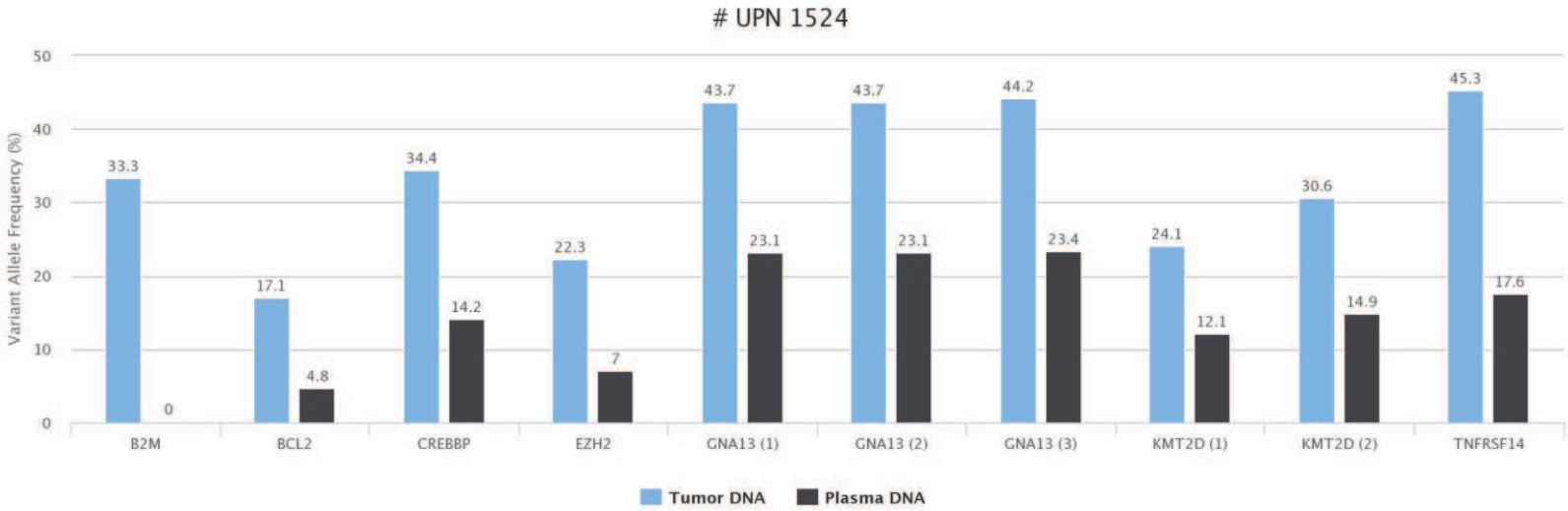
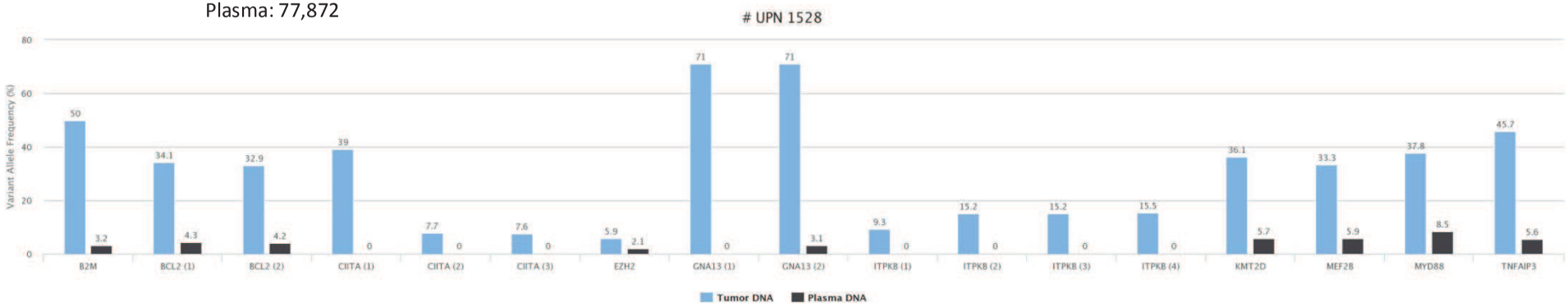


Figure S1

Relevant reads sequenced
Tumor: 3421
Plasma: 77,872



Relevant reads sequenced
Tumor: 636
Plasma: 386

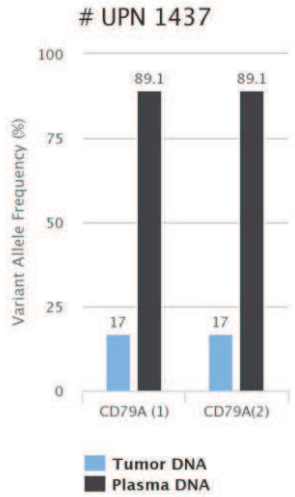
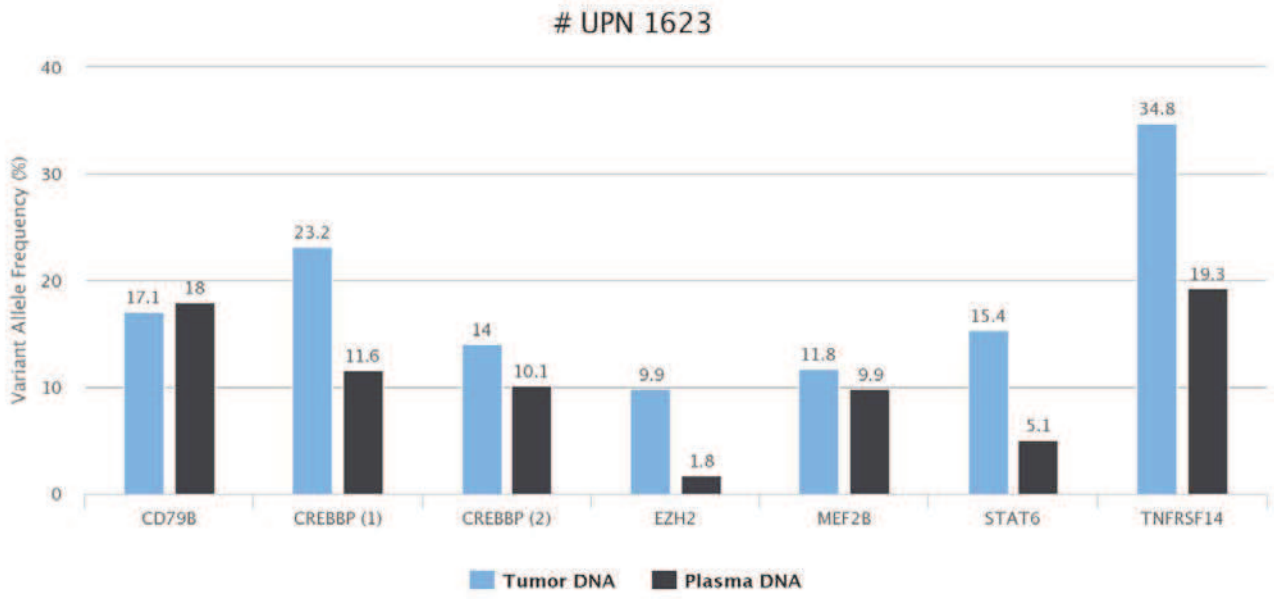


Figure S1

Relevant reads sequenced
Tumor: 3,724
Plasma: 36,037



Relevant reads sequenced
Tumor: 1,811
Plasma: 58,131

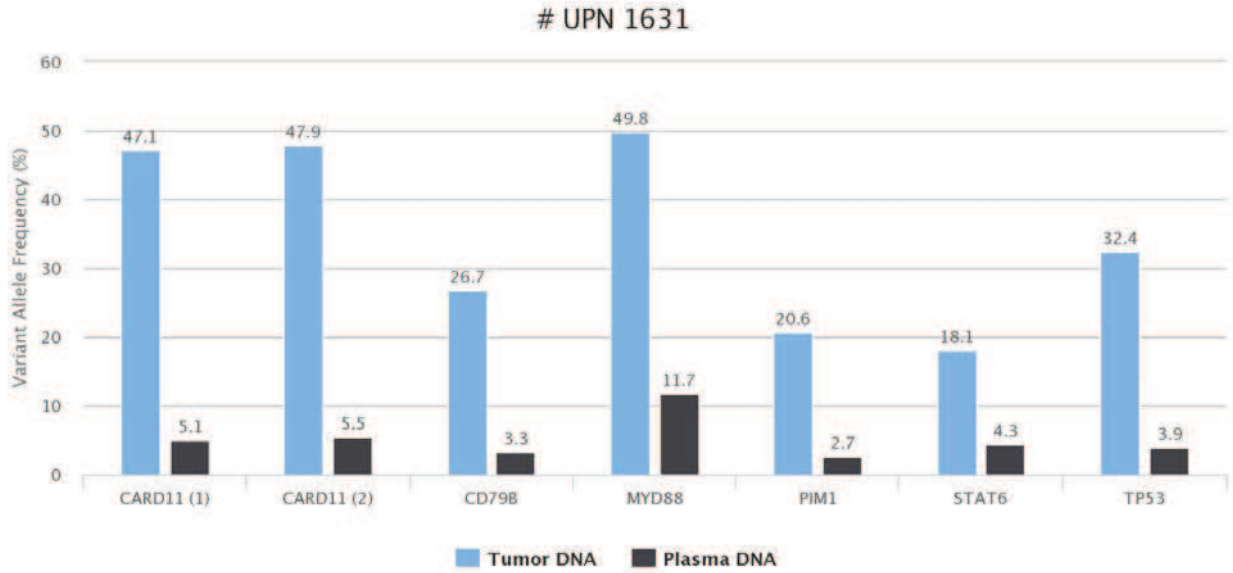
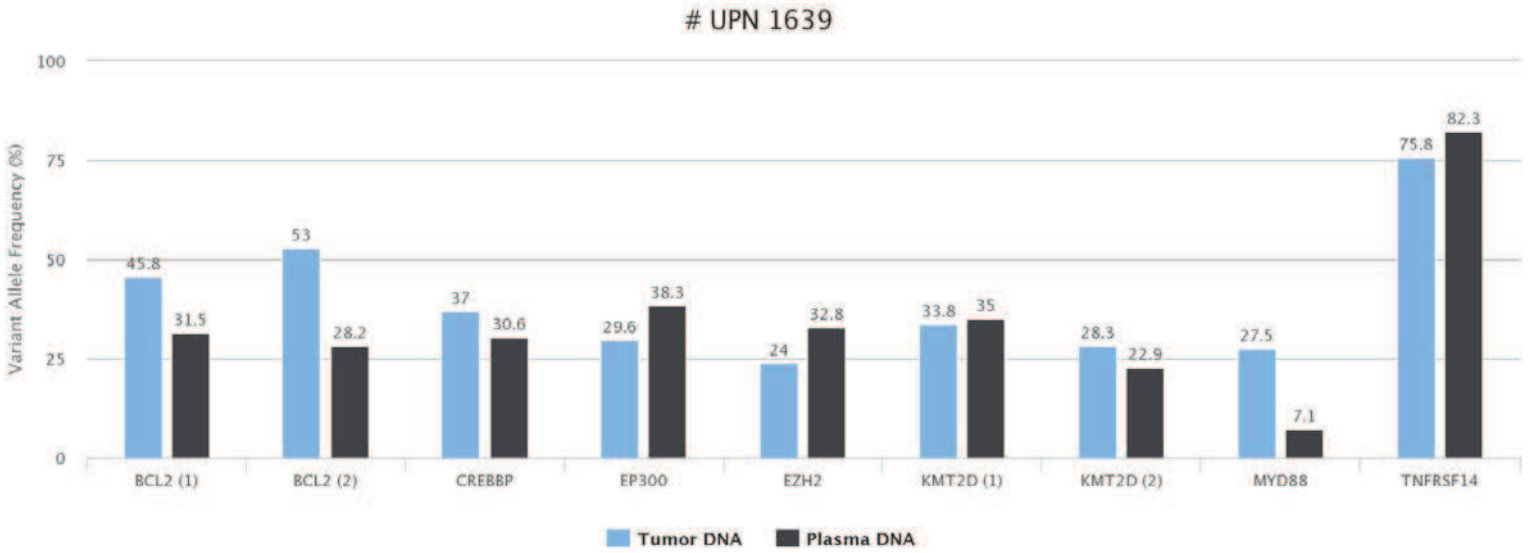
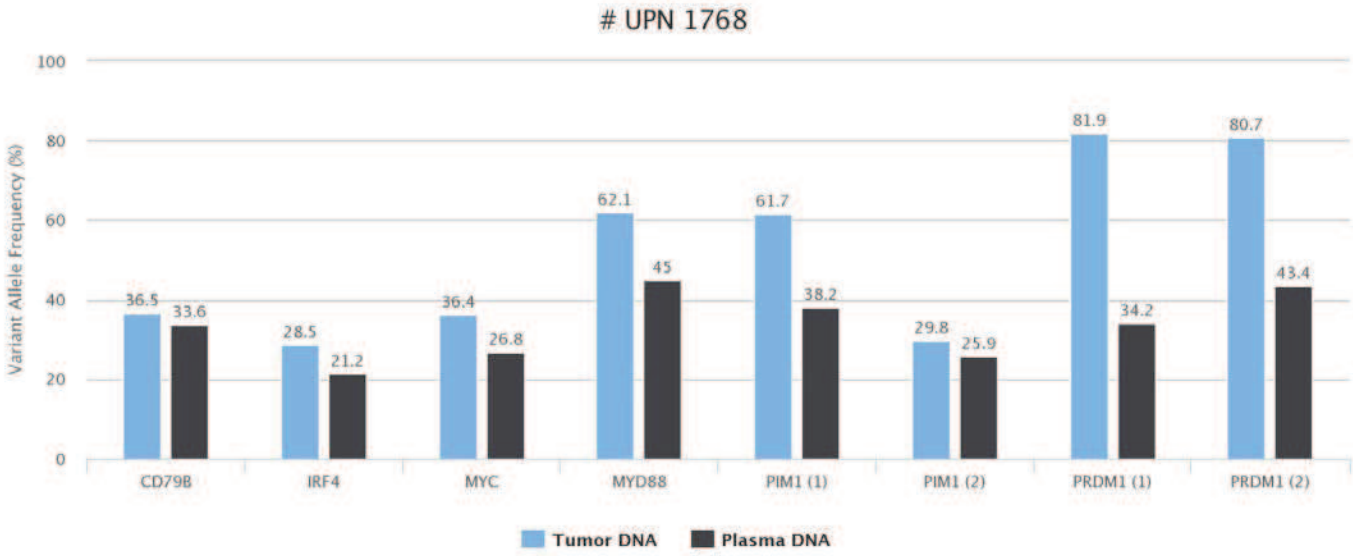


Figure S1

Relevant reads sequenced
Tumor: 2,041
Plasma: 85,369



Relevant reads sequenced
Tumor: 1,852
Plasma: 50,862



Supplementary table 1

UPN	Sex	Age	Stage	IPI	LDH (x UNL)	Bone marrow involvement	Phenotype	Gene	SNV	Tumor DNA			circulating DNA					Duration of Tumor/plasma storage (y) before sequencing
										VAF (%)	read number	mean VAF (%)	VAF (%)	read number	mean VAF (%)	Allele Call	concentration (ng/µl)	
964	F	17	IV	3	5	0	GCB	*ITPKB*	NM_002221:exon2:c.C731T-(p.T244I)	40.9	79/193	40.1	2.5	158/6313	2.3	Heterozygous	2.02	14
								ITPKB	NM_002221:exon2:c.726_728del-(p.242_243del)	40.6	78/192		2.4	152/6260		Heterozygous		
								CARD11	NM_032415:exon7:c.G1010A-(p.R337Q)	17.1	13/76		0.0	0/54		Absent		
								EZH2	NM_004456:exon16:c.A1937T-(p.Y646F)	26.3	76/289		2.8	126/4435		Heterozygous		
								B2M	NM_004048:exon1:c.T2A-(p.M1K)	75.6	102/135		4.0	130/4016		Heterozygous		
1003	F	75	III	3	2.5	0	ABC	*ITPKB*	NM_002221:exon2:c.541delC-(p.R181fs)	35.3	36/102	40.0	1.4	22/1540	5.1	Heterozygous	1.94	13
								MYD88	NM_002468:exon4:c.T695C-(p.M232T)	26.5	67/253		0.0	0/21		Absent		
								CIITA	NM_000246:exon13:c.C2819A-(p.T940K)	51.9	14/27		16.7	1/6		Heterozygous		
								CD79B	NM_000626:exon6:c.596delT-(p.L199fs)	46.3	93/201		2.1	369/17260		Heterozygous		
1251	M	42	IE	0	0.8	0	GCB	*TNFRSF14*	NM_003820:exon1:c.G3A-(p.M1I)	52.5	74/141	30.9	0.0	0/1113	0.0	Absent	0.46	10
								EZH2	NM_004456:exon16:c.A1937T-(p.Y646F)	9.8	38/389		0.0	0/2519		Absent		
								SOCS1	NM_003745:exon2:c.C174A-(p.F58L)	30.4	7/23		0.0	0/1053		Absent		
1437	M	76	III	4	5.85	0	ABC	*CD79A*	NM_001783:c.568-2A>G	17.0	53/318	17.0	89.1	344/386	89.1	Heterozygous	11.2	8
								CD79A	NM_001783:exon5:c573_617del-(p.191_206del)	17.0	53/318		89.1	344/386		Heterozygous		
1524	F	45	III	2	1.52	0	GCB	*TNFRSF14*	NM_003820:exon3:c.T269C-(p.L90P)	45.3	77/170	33.9	17.6	539/3064	14.0	Heterozygous	1.48	7
								EZH2	NM_004456:exon16:c.T1936A-(p.Y646N)	22.3	106/475		7.0	818/11751		Heterozygous		
								KMT2D	NM_003482:exon40:c.G13671C-(p.Q4557H)	24.1	64/266		12.1	1269/10509		Heterozygous		
								KMT2D	NM_003482:exon34:c.C8401T-(p.R2801X)	30.6	33/108		14.9	1015/6818		Heterozygous		
								B2M	NM_004048:exon1:c.T35C-(p.L12P)	33.3	44/132		0.0	0/4555		Absent		
								CREBBP	NM_004380:c.4394+2A>T	34.4	45/131		14.2	47/331		Heterozygous		
								GNA13	NM_006572:exon3:c.T533A-(p.L178Q)	43.7	52/119		23.1	1909/8254		Heterozygous		
								GNA13	NM_006572:exon3:c.T530C-(p.F177S)	43.7	52/120		23.1	1920/8314		Heterozygous		
								GNA13	NM_006572:exon3:c.T526A-(p.Y176N)	44.2	53/120		23.4	1946/8312		Heterozygous		
								BCL2	NM_000633:exon2:c.G17C-(p.R6T)	17.1	42/245		4.8	409/8558		Heterozygous		
1528	F	66	III	4	1.39	0	GCB	*ITPKB*	NM_002221:exon2:c.C1583T-(p.S528L)	9.3	14/151	31.0	0.0	0/1713	2.5	Absent	1.94	7
								ITPKB	NM_002221:exon2:c.C1471G-(p.P491A)	15.2	19/125		0.0	0/3582		Absent		
								ITPKB	NM_002221:exon2:c.C1454A-(p.A485E)	15.2	19/125		0.0	0/3577		Absent		
								ITPKB	NM_002221:exon2:c.C1438G-(p.P480A)	15.5	19/123		0.0	0/3570		Absent		
								MYD88	NM_002468:exon4:c.G728A-(p.S243N)	37.8	74/196		8.5	161/1888		Heterozygous		
								TNFAIP3	NM_006290:exon9:c.T2335C-(p.C779R)	45.7	32/70		5.6	177/3152		Heterozygous		
								EZH2	NM_004456:exon16:c.T1936A-(p.Y646N)	5.9	44/741		2.1	152/7401		Heterozygous		
								KMT2D	NM_003482:exon48:c.C15142T-(p.R5048C)	36.1	122/338		5.7	207/3620		Heterozygous		
								B2M	NM_004048:exon1:c.G3A-(p.M1I)	50.0	63/126		3.2	78/2465		Heterozygous		
								CIITA	NM_000246:exon11:c.2239delT-(p.Y747fs)	39.0	30/77		0.0	0/2130		Absent		
								CIITA	NM_000246:exon18:c.T3236A-(p.V1079D)	7.7	14/183		0.0	0/9150		Absent		
								CIITA	NM_000246:exon18:c.T3241A-(p.Y1081N)	7.6	14/184		0.0	0/9162		Absent		
								GNA13	NM_006572:exon1:c.G201C-(p.Q67H)	71.0	105/148		0.0	0/2695		Absent		
								GNA13	NM_006572:exon1:c.T158C-(p.L53P)	71.0	105/148		3.1	82/2672		Heterozygous		
								BCL2	NM_000657:exon2:c.G589A-(p.G197S)	34.1	47/138		4.3	214/4929		Heterozygous		
								BCL2	NM_000633:exon2:c.C502G-(p.P168A)	32.9	45/137		4.2	209/5013		Heterozygous		
								MEF2B	NM_001145785:exon3:c.G229C-(p.E77Q)	33.3	137/411		5.9	654/11153		Heterozygous		

Supplementary table 1

1559	F	83	IE	1	0.8	0	ABC	*CD58*	NM_001779:exon3:c.C454T-(p.R152X)	92.9	92/99	64.0	0.0	0/13231	<1	Absent	0.618	7
								MYD88	NM_002468:exon5:c.T778C-(p.L265P)	51.8	187/361		0.0	0/11427		Absent		
								CREBBP	NM_004380:exon30:c.G5105T-(p.R1702L)	46.3	154/333		0.0	0/5663		Absent		
								TP53	NM_000546:exon5:c.G524A-(p.R175H)	89.0	105/118		0.0	0/13012		Absent		
								CD79A	NM_001783:c.568-2A>G	52.0	122/235		0.5	20/3930		Absent		
								CD79A	NM_001783:exon5:c573_617del-(p.191_206del)	52.0	122/235		0.5	20/3932		Absent		
1586	F	62	III	2	0.85	0	NA	*TNFRSF14*	NM_003820:exon2:c.C164T-(p.P55L)	13.8	44/318	20	2.1	6/284	<1	Heterozygous	1.31	6
								EZH2	NM_004456:exon16:c.A1937T-(p.Y646F)	13.8	41/298		0.0	0/6575		Absent		
								STAT6	NM_001178078:exon12:c.A1256G-(p.D419G)	18.9	83/439		0.0	2/5517		Absent		
								CREBBP	NM_004380:exon26:c.G4283C-(p.R1428P)	33.6	49/146		1.3	65/4894		Absent		
								GNA13	NM_006572:c.561+2A>G	17.7	11/62		0.0	0/4391		Absent		
1623	M	53	IV	2	1.03	0	GCB	*TNFRSF14*	NM_003820:c.179-1G>C	34.8	146/419	18	19.3	694/3587	11	Heterozygous	0.95	6
								EZH2	NM_004456:exon16:c.A1937T-(p.Y646F)	9.9	57/575		1.8	62/3480		Heterozygous		
								STAT6	NM_001178078:exon12:c.A1256G-(p.D419G)	15.4	79/514		5.1	200/3889		Heterozygous		
								CREBBP	NM_004380:exon28:c.A4628T-(p.D1543V)	23.2	166/717		11.6	489/4222		Heterozygous		
								CREBBP	NM_004380:exon27:c.T4504C-(p.W1502R)	14.0	54/387		10.1	328/3239		Heterozygous		
								CD79B	NM_000626:exon6:c.611dupC-(p.T204fs)	17.1	83/486		18.0	839/4651		Heterozygous		
								MEF2B	NM_001145785:exon3:c.G229A-(p.E77K)	11.8	74/626		9.9	1283/12969		Heterozygous		
1631	M	64	I	2	1.54	0	ABC	*MYD88*	NM_002468:exon4:c.G728A-(p.S243K)	49.8	220/442	35	11.7	655/5586	5	Heterozygous	1.1	6
								PIM1	NM_002648:exon4:c.G290A-(p.S97N)	20.6	14/68		2.7	130/4797		Heterozygous		
								CARD11	NM_032415:exon5:c.645_647del-(p.215_216del)	47.1	99/210		5.1	154/3028		Heterozygous		
								CARD11	NM_032415:exon4:c.T343C-(p.F115L)	47.9	68/142		5.5	344/6300		Heterozygous		
								STAT6	NM_001178078:exon12:c.A1249G-(p.N417D)	18.1	96/532		4.3	570/13250		Heterozygous		
								TP53	NM_000546:exon5:c.T518A-(p.V173E)	32.4	72/222		3.9	522/13317		Heterozygous		
								CD79B	NM_000626:exon5:c.A587T-(p.Y196F)	26.7	52/195		3.3	386/11853		Heterozygous		
1639	M	69	IV	2	2.4	0	GCB	*TNFRSF14*	NM_003820:exon6:c.C620A-(p.S207X)	75.8	47/62	39	82.3	2400/2915	34	Heterozygous	10.2	6
								MYD88	NM_002468:exon3:c.C840G-(p.S219C)	27.5	133/483		7.1	1616/22541		Heterozygous		
								EZH2	M_004456:exon16:c.A1937T-(p.Y646F)	24.0	126/526		32.8	3259/9939		Heterozygous		
								KMT2D	NM_003482:exon34:c.C9838T-(p.Q3280X)	33.8	187/554		35.0	5878/16784		Heterozygous		
								KMT2D	NM_003482:exon22:c.5296_5297insGAGG-(p.D1766fs)	28.3	13/46		22.9	671/2929		Heterozygous		
								CREBBP	NM_004380:c.4560+2A>G	37.0	17/46		30.6	2918/9530		Heterozygous		
								BCL2	NM_000633:exon2:c.C176A-(p.P59Q)	45.8	11/24		31.5	223/709		Heterozygous		
								BCL2	NM_000633:exon2:C20T-(p.T7I)	53.0	98/185		28.2	2267/8032		Heterozygous		
								EP300	NM_001429:exon27:c.4373_4375del-(p.1458_1459del)	29.6	34/115		38.3	4597/11990		Heterozygous		
1768	M	83	IV	3	4.2	0	ABC	*MYD88*	NM_002468:exon5:c.T778C-(p.L265P)	62.1	216/348	52	45.0	5936/13194	34	Heterozygous	1.82	4
								IRF4	NM_002460:exon2:c.G53A-(p.S18N)	28.5	75/263		21.2	929/4374		Heterozygous		
								PIM1	NM_002648:exon3:c.237delG-(p.E79fs)	61.7	145/235		38.2	1541/4030		Heterozygous		
								PIM1	NM_002648:exon4:c.G382C-(p.D128H)	29.8	34/114		25.9	987/3812		Heterozygous		
								PRDM1	NM_001198:exon2:c.G284A-(p.S95N)	81.9	149/182		34.2	1273/3724		Heterozygous		
								PRDM1	NM_001198:exon5:c.1103_1112del-(p.368_371del)	80.7	221/274		43.4	5076/11689		Heterozygous		
								MYC	NM_002467:exon2:c.C145T-(p.Q49X)	36.4	102/280		26.8	262/979		Heterozygous		
								CD79B	NM_000626:exon5:c.A587C-(p.Y196S)	36.5	57/156		33.6	3041/9060		Heterozygous		

Supplementary Table 1: Clinical characteristics and list of somatic variants (insertion/deletion/ single nucleotide variant) detected by sequencing in tumor DNA and cell-free plasma circulating DNA.

ABC: Activated B-Cell like; GCB: Germinal Center B-cell like; In/del: insertion/deletion; LDH: Lactate dehydrogenase; IPI: International Prognosis Index; SNV: single nucleotide variant; UPN: Unique Personal Number; ULN: upper limit value; VAF: variant allele frequency